



SNHU Academic Archive

SNHU Academic Archive search configuration

Last modification date: December 2009

This document explains the current search configuration for the SNHU Academic Archive, and explains the reasons behind the current configuration.

When searching DSpace:

- The Boolean term ***and*** is used by default. In other words, all search terms must be present to achieve a result.
- Search results are returned from elements indexed below and the full text of documents in the Academic Archive.

The DSpace search is configured to only index certain metadata elements, as noted in the tables below. That means that if a particular metadata element is not in the list of indexed elements, it will not be searchable.

For more information about what type of metadata each element contains, as well as information about an element's qualifiers, please refer to the SNHU Academic Archive Data Dictionary, <http://hdl.handle.net/10474/14>

Elements that are indexed by DSpace

An asterisk means that all qualifiers to that element are indexed.

Element Name	Reason for Indexing
contributor.author contributor.editor creator	Authors or contributing authors of a document are commonly queried elements.
contributor.advisor contributor.committeeMember	Useful for researchers interested in works associated with a particular professor. Recommended by Hyun Hee Kim and Yong Ho Kim (2008) in their "Usability Study of Digital Institutional Repositories" doi: 10.1108/02640470810921637
date.issued date.copyright	Relevant to researchers searching for a particular work, or works from a specific time period.
description.abstract description.tableOfContents description.degree description.school description.program description subject*	Supply descriptive information facilitating keyword search results.

title*	Titles of documents are commonly queried elements.
type	Type, controlled by the <i>MARC Value List for Genre Terms</i> , is useful for searchers who may keyword “Jacobs thesis” when searching for a thesis by the author Jacobs.

Elements that are not indexed by DSpace

Some of these elements do not necessarily have the potential to harm a search, but there is no particular reason to include them, either. Since indexing takes longer as more elements are included, it is judged best to leave them out.

Element Name	Reason for not indexing
contributor contributor.other	Including results from other contributors can complicate results when a search is conducted for a primary author.
date.accessioned date.available	When searching for a particular date, the user expects to see the copyright or issue date of a document in the results; the accessioned date or date available is meaningless to these users. Results including date.accessioned and date.available caused problems for users in the usability study by Jihyun Kim (http://eprints.rclis.org/5189/ , page 7).
identifier*	Currently the only identifier is the Handle System URL. Since this identifier in DSpace is a complete URL, it can simply be entered into the web address field of any browser. Including it in a search index might complicate other searches that involve numbers, such as dates. No other types of identifiers are currently used.
description.provenance	This element is created automatically by DSpace to describe preservation information such as who added the document, when, and includes bitstream and checksum information. Names included in the provenance data may also be authors in the repository. If this field were indexed, a keyword search would result in all of the items approved for entry by that person.

format.extent	Administrative information not likely to be searched by end-users.
language.iso	Languages are stored as an ISO abbreviation, which is not likely to be searched by a user.
publisher	Currently, everything in the repository will be cataloged as published by Southern New Hampshire University. A search would result in all the items in the archive and therefore not be useful.
relation.requires	Administrative information not likely to be searched by end-users.
rights	This is a blanket rights statement common among all records under copyright, so would not be useful for a search.
source	Administrative information not likely to be searched by end-users.
description.bibliographicCitation	Duplicates date, title, author information.
digSpecs	Administrative information not likely to be searched by end-users.
format.mediaType	Administrative information not likely to be searched by end-users.
rightsHolder	Generally the rights holder is also the author or publisher, and would therefore not be useful for a search.

These search indexes are implemented in the “dspace.cfg” file in the config directory, DSpace 1.5.2, under the “Fields to Index for Search” subheading.